CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION CFSTI
DOCUMENT MANAGEMENT BRANCH 410.11

LIMITATIONS IN REPRODUCTION QUALITY

ACCESSION # AD 604 239

☒ 1. WE REGRET THAT LEGIBILITY OF THIS DOCUMENT IS IN PART UNSATISFACTORY. REPRODUCTION HAS BEEN MADE FROM BEST AVAILABLE COPY.

□ 2. A PORTION OF THE ORIGINAL DOCUMENT CONTAINS FINE DETAIL WHICH MAY MAKE READING OF PHOTOCOPY DIFFICULT.

□ 3. THE ORIGINAL DOCUMENT CONTAINS COLOR, BUT DISTRIBUTION COPIES ARE AVAILABLE IN BLACK-AND-WHITE REPRODUCTION ONLY.

□ 4. THE INITIAL DISTRIBUTION COPIES CONTAIN COLOR WHICH WILL BE SHOWN IN BLACK-AND-WHITE WHEN IT IS NECESSARY TO REPRINT.

□ 5. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED, DOCUMENT WILL BE AVAILABLE IN MICROFICHE ONLY.

□ 6. LIMITED SUPPLY ON HAND: WHEN EXHAUSTED DOCUMENT WILL NOT BE AVAILABLE.

□ 7. DOCUMENT IS AVAILABLE IN MICROFICHE ONLY.

□ 8. DOCUMENT AVAILABLE ON LOAN FROM CFSTI (TT DOCUMENTS ONLY).

□ 9.

NBS 9/64

PROCESSOR:

ON THE CONTINUOUS GOLD-MINING EQUATION

Richard Bellman and Sherman Lehman

✓ P-436 ✓ Bel

1 October 1953

The RAND Corporation

1700 MAIN ST. • SANTA MONICA • CALIFORNIA

ON THE CONTINUOUS GOLD-MINING EQUATION

Richard Bellman and Sherman Lehman
The RAND Corporation and Stanford University

§1. Introduction.

In some previous communications, [1] and [2], we have described some results obtained in the investigation of a dynamic programming problem, the "gold-mining" problem, which led to the functional equation

$$f(x,y) = \text{Max} \begin{bmatrix} A: & \sum_{i=1}^{N} p_i(r_i x + f((1-r_i)x,y) \\ B: & \sum_{i=1}^{N} q_i(s_i y + f(x,(1-s_i)y) \end{bmatrix}, \quad x,y \geq 0 \qquad (1.1)$$

where p_i, $q_i \geq 0$, $\sum_i p_i$, $\sum_i q_i < 1$. The solution of this equation was given, inter alia, in [1] and shown to have a relatively simple form. In addition, a partial solution of a more complicated equation, corresponding to a nonlinear utility function, was given in [2], having the same form. It is known, however, as a result of an unpublished counter-example due to H. N. Shapiro and S. Karlin, that the solution of more general equations such as

$$f(x,y) = \text{Max} \begin{bmatrix} A: & p_1(r_1 x + f((1-r_1)x,y) \\ B: & q_1(r_2 y + f(x,(1-r_2)y) \\ C: & p_2(r_2 x + r_4 y + f((1-r_3)x, (1-r_4)y) \end{bmatrix} \qquad (1.2)$$

$0 \leq r_1,r_2,r_3,r_4 \leq 1$, $0 < p_1,q_1,p_2 < 1$ cannot have the same simple form for all values of the parameters.

In an effort to gain some insight into the structure of particular classes of solutions of (1.2), and similar equations of more complicated type, we have been led to consider some continuous analogues of these equations. There are many different procedures for obtaining these continuous analogues. One which we have followed to begin with leads to problems in the calculus of variations.

An essential feature of our research lies in viewing a policy in its extensive rather than normal form, to borrow the terminology of game theory. Another way of stating this is that instead of determining the complete solution for one set of initial parameters, which would correspond to determining the extremal curve in the classical theory of the calculus of variations, we attack our problem by imbedding it into the family of problems of this type with arbitrary initial parameters. This is the approach used throughout the theory of dynamic programming, cf. [1], [2], [3], [4], [5]. Having done this we determine an optimal continuation from each position, which upon being carried through yields an optimal policy.

This approach, which may be considered a variant in problems of deterministic type, is in many ways a necessity in problems of stochastic type. It is possible to treat many of the classical problems in the calculus of variations by means of this technique. We shall return to this point at some future time.

Guided by our knowledge of the solution in the discrete cases, and using the behavioristic approach described above, we have been able to solve completely and explicitly a variety of problems which are intractable in the original discrete form.

In the following sections we shall discuss the simplest counterpart of (1.2) in continuous form, and list a number of typical results we have obtained. Following this we shall sketch briefly a formulation of the more general continuous version which results from processes corresponding to (1.1) and which requires more powerful techniques.

A more complete discussion and proofs of the results contained herein will appear elsewhere. Further results concerning more general problems discussed in [3] will be presented subsequently.

§2. Formulation.

In the formulation of problems involving the use of continuous policies we are immediately faced with the difficulty of defining what we mean by a continuous mixed strategy, and of constructing the appropriate mathematical theory with which to handle this therny concept. To circumvent these conceptual and mathematical difficulties, we shall utilize an idea emphasized in [6], which—briefly put—is that for mathematical purposes, mixing at a point is to an arbitrary degree of approximation equivalent to mixing pure strategies in an interval about the point.

Let us then consider a process where we are given two initial quantities, x and y, the gold mines of [2]; and two operations, A and B, mining operations. If A is used over a time interval δ, there is a probability $1 - q_1\delta + o(\delta)$ that $r_1x\delta + o(\delta)$ is obtained and that the process is allowed to continue, with the new initial amounts $x - r_1x\delta + o(\delta), y$; and a probability $q_1\delta + o(\delta)$ that nothing is obtained and the process terminates. In a like manner, if B is used, there is a probability $1 - q_2\delta + o(\delta)$ that $r_2y\delta + o(\delta)$ is obtained and the process continues; and a probability $q_2\delta + o(\delta)$ that the process terminates.

To introduce the concept of mixing, we consider first the case in which the time interval is divided into intervals of length Δ, where Δ is small. In a typical interval $[t, t+\Delta]$, $t = k\Delta$, the first part of the interval, $[t, t+\phi_1\Delta]$, will be devoted to the use of A; while the second part, $[t+\phi_1\Delta, t+\Delta]$, will be devoted to the use of B. In the limit, as $\Delta \longrightarrow 0$, we obtain the effect of mixing A and B at t in the ratio $\phi_1 : (1-\phi_1)$, cf. [6] for further discussion.

A strategy consists of a choice of ϕ_1 for each of the points $k\Delta$. We wish to determine the strategy which will maximize the expected value of the amount obtained before the process terminates. For any given strategy let

$$
\begin{aligned}
x(t) &= \text{quantity of gold remaining in first mine provided that the process has continued until } t,\\
y(t) &= \text{quantity of gold remaining in second mine provided that the process has continued until } t,\\
p(t) &= \text{probability that the process continues at least to } t,\\
f(t) &= \text{expected amount obtained up to } t.
\end{aligned}
\tag{2.1}
$$

Writing down the equations expressing $x(t+\Delta)$, $y(t+\Delta)$, $p(t+\Delta)$, $f(t+\Delta)$ in terms of the values at t, and letting $\Delta \longrightarrow 0$, we are led to the following system of differential equations:

$$
\begin{aligned}
\frac{dx}{dt} &= -\phi_1(t)r_1x(t), & x(0) &= x_0,\\
\frac{dy}{dt} &= -\phi_2(t)r_2y(t), & y(0) &= y_0,\\
\frac{dp}{dt} &= -p(t)\left[\phi_1(t)q_1 + \phi_2(t)q_2\right], & p(0) &= 1,\\
\frac{df}{dt} &= p(t)\left[\phi_1(t)r_1x(t) + \phi_2(t)r_2y(t)\right], & f(0) &= 0,
\end{aligned}
\tag{2.2}
$$

where $0 \leq \phi_1 \leq 1$, $\phi_2 = 1 - \phi_1$. The problem is now to determine $\phi_1(t)$ so as to maximize $f(\infty)$. It is not difficult to give a proof based upon, say, weak convergence, which will assure us that the maximum is actually attained. As W. Fleming has kindly informed us, the existence of a maximum is guaranteed by a general theorem in the calculus of variations.

Since the equations are fortunately nonlinear, variational techniques are particularly applicable. We find

Theorem 1. *The maximum value of* $f(\infty)$ *is attained by the policy*

$$
\begin{aligned}
&\text{(a) } \textit{If } q_2r_1x > q_1r_2y, \quad \phi_1 = 1,\\
&\text{(b) } \textit{If } q_1r_2y > q_2r_1x, \quad \phi_2 = 1,\\
&\text{(c) } \textit{If } q_2r_1x = q_1r_2y, \quad \phi_1 = r_2/(r_1+r_2), \ \phi_2 = r_1/(r_1+r_2).
\end{aligned}
\tag{2.3}
$$

Note that the boundary line is, as might be expected, the set of points where expected gain over expected cost is the same for both choices, A and B.

Theorem 2. *If* T *is finite, the optimal policy has one of the following six forms:*

$$\begin{array}{ll} \text{(a) A } \textit{always}, & \text{(d) A, } \textit{then} \text{ M, } \textit{then} \text{ A,} \\ \text{(b) B } \textit{always}, & \text{(e) B, } \textit{then} \text{ M, } \textit{then} \text{ A,} \\ \text{(c) M } \textit{followed by} \text{ A} & \text{(f) B } \textit{followed by} \text{ A.} \end{array} \tag{2.4}$$

This is for $q_1 < q_2$; *a similar result holds for* $q_2 < q_1$. *The precise intervals within which each is used may be determined explicitly*. Here M represents the choice given in (2.3c).

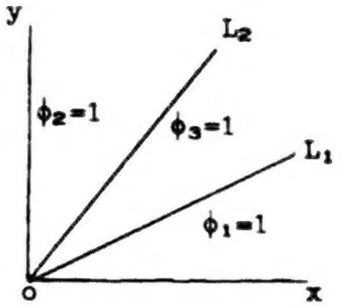
The optimal strategy represents a compromise between the long-term policy given in Theorem 1 and the short-term policy of maximizing expected gain.

The 3-choice problem corresponding to (1.2) has the continuous analogue:

$$\begin{aligned} \frac{dx}{dt} &= -\left[\phi_1(t)r_1 + \phi_3(t)r_3\right]x(t), \\ \frac{dy}{dt} &= -\left[\phi_2(t)r_2 + \phi_3(t)r_4\right]y(t), \\ \frac{dp}{dt} &= -p(t)\left[\phi_1(t)q_1 + \phi_2(t)q_2 + \phi_3(t)q_3\right], \\ \frac{df}{dt} &= p(t)\left[(\phi_1(t)r_1 + \phi_3(t)r_3)x(t) + (\phi_2(t)r_2 + \phi_3(t)r_4)y(t)\right], \\ & \qquad f(o) = 0, \end{aligned} \tag{2.5}$$

where $0 \leq \phi_1, \phi_2, \phi_3 \leq 1$, $\phi_1 + \phi_2 + \phi_3 = 1$.

The maximum value of $f(oo)$ is provided in the general case by the policy represented schematically by



(2.6)

Depending upon the values of the parameters, one line is an absorbing barrier, which is to say, for (x,y) on the line a mixed policy is pursued which keeps the point on the line. This line has as its equation the equality of expected gain over expected cost. The other line will be a translucent barrier, causing only a change from $\phi_i = 1$ to $\phi_j = 1$, and is not defined by an equality of the above type. In special cases the middle region disappears and L_1 coincides with L_2. The solution is now that for the two-choice case.

This last result is quite surprising and explains some of the difficulties of the discrete problem. One boundary, the absorbing barrier, is determined by a local condition, whereas the other is determined by a global condition.

Theorem 3. *If, in the two-choice problem described by* (2.2), *in place of expected return, we seek to maximize the expected value of some function ϕ of the total return where ϕ is any strictly increasing function, the solution is that given by* (2.3).

To obtain this result we consider

$$G = - \int_0^\infty \phi(x_0 + y_0 - x(t) - y(t))dp(t), \qquad (2.7)$$

the quantities being defined as in (2).

The proofs of the above results are long and detailed, depending upon a precise analysis of the properties of an optimal policy.

§3. *More General Processes*.

Let us now consider the more general process corresponding to (1.1). Here the use of A leads to a variety of possible gains, and similarly for B. The quantities $x(t)$ and $y(t)$, as defined by (2.1) are now stochastic quantities. This means that it is no longer possible to obtain the equations of (2.2). Instead we must introduce the function $F(u,v,t)$ defined by the property that

$$\Pr\left\{u \leq x(t) \leq u + du,\ v \leq y(t) \leq v + dv\right\} = F(u,v,t)dudv. \quad (3.1)$$

We may now, in a way similar to that followed in o2, derive a partial differential equation for F of the form

$$\frac{\partial F}{\partial t} = P(u,v)\frac{\partial F}{\partial u} + Q(u,v)\frac{\partial F}{\partial v}. \quad (3.2)$$

The system of ordinary differential equations

$$\frac{du}{dt} = P(u,v), \quad \frac{dv}{dt} = Q(u,v) \quad (3.3)$$

connected with (3.2) will have a form similar to the first two equations in (2.2).

The differential equations we have used to define our continuous processes bear the same relation to the rigorous integral equations defined by the original processes as the heat equations bears to the Chapman-Kolmogoroff equations.

Finally, let us note that the above formalism is also applicable to two-person multi-stage games of continuous type, and, in particular, to pursuit games.

These extensions will be discussed in subsequent communications.

BIBLIOGRAPHY

1. Bellman, R. "On the Theory of Dynamic Programming," Proc. Nat. Acad. Sci., 38 (1952), pp. 716-719.

2. ————. "Some Functional Equations in the Theory of Dynamic Programming," Proc. Nat. Acad. Sci. (to appear).

3. ————. "Bottleneck Problems and the Theory of Dynamic Programming," Proc. Nat. Acad. Sci. (to appear).

4. ————. "A Problem in the Theory of Dynamic Programming," Econometrica (to appear).

5. ————. "Computational Problems in the Theory of Dynamic Programming," Proc. of Symposium on Numerical T Analysis, Santa Monica, 1953.

6. Bellman, R. and Blackwell, D. "Some Two-person Games Involving Bluffing," Proc. Nat. Acad. Sci., 35 (1949), pp. 600-605.

bjc